Mammalian Genome 7, 657–666 (1996).

Mammalian Genome

A survey of the goat genome transcribed in the lactating mammary gland

F. Le Provost, A. Lépingle, P. Martin

Laboratoire de Génétique Biochimique et de Cytogénétique, Institut National de la Recherche Agronomique, 78352 Jouy-en-Josas Cedex, France

Received: 25 January 1996 / Accepted: 25 April 1996

Abstract. To fulfill its primary function, which is to synthesize milk during the course of lactation, the mammary gland requires efficient transcriptional, translational, and secretory machineries involving multiple genes among which promising candidates underlying the genetic variation of milk production have to be found. With the aim of providing a first transcriptional profile of lactating mammary tissue, a non-normalized cDNA library has been constructed from the udder of a lactating goat. After having discarded cDNA clones encoding the major milk proteins the rapid characterization of genes expressed in this tissue, by automated partial cDNA sequencing, was used to analyze a total of 435 cDNA clones. Examination of the Expressed Sequence Tags (ESTs) for similarities with sequence databases identified 234 cDNAs corresponding to 140 unique genes or proteins. Eighty-three clones, not similar to any current database entries, representing 77 novel sequences unrelated to previously described genes, were thus identified. Tissue specificity and relative abundance of 18 of these 77 unidentified clones were examined by dot blot and RT-PCR experiments. Sequence data were subsequently used to assign six genes of unknown localization in the bovine genome, to synteny groups by use of bovine–hamster cell hybrids and PCR.

Introduction

The basic function of the lactating mammary gland is to provide the mammalian neonate with milk that contains a balance of nutrients for growth and development. This key nutritional role and thus commercial value triggered numerous efforts that have been directed to understanding the development of the mammary gland, particularly in terms of cellular proliferation, differentiation and lactogenesis. During lactation, mammary epithelial cells synthesize and secrete large quantities of specific milk components including proteins, lipids, and lactose. This process involves the activation of ubiquitously expressed "housekeeping" genes as well as genes specifically transactivated in mammary epithelial cells. Many intimate mechanisms of the mammary function remain to be deciphered, particularly some of the most basic processes involved in milk synthesis and secretion. A main goal is to identify genes controlling genetic variation of this function in order to use their polymorphisms to improve milk production traits.

Currently, genetic improvement is based on quantitative genetic methods, by use of data for a restricted number of traits (primarily milk yield, protein and fat contents). In addition, qualitative traits, which are also under more or less strict genetic control, deserve to receive special attention and to be subjected to genetic improvement (Gellin and Grosclaude 1991). Improvements in selection methods (Stranzinger 1987) might be achieved by identifying regions of the genome occurring in the variability of Economic Trait Loci (ETL; Geldermann 1975; Beckmann and Soller 1988). Such an approach is strongly dependent on the isolation of multi-allelic genetic markers and coding sequences that are needed to construct an integrated meiotic and cytogenetic map of genomes. For the bovine genome, genetic linkage maps have already been published, and 877 loci, including 563 anonymous loci and 314 genes or pseudogenes have been assigned (Eggen and Fries 1995). The power of this approach was recently exemplified by Georges et al. (1995), who used a battery of bovine microsatellite markers to map quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing.

An alternative strategy is direct identification of genes involved in the determination of milk production traits that are expressed in the mammary gland. These genes should be subsequently characterized with the aim of finding possible polymorphisms associated with some variability of these traits. In this respect, the α_{s1}-casein model provides an illustrative example, since the genetic polymorphisms described at this locus may account for an increase (up to 20%) in the protein content of goat milk (reviewed in Grosclaude et al. 1994). A rapid and efficient way to identify such genes (known and/or unknown) is to characterize their transcripts by performing a systematic partial sequencing of cDNA library clones, thus generating Expressed Sequence Tags (ESTs) derived from the mammary gland. This approach has been used to establish a detailed profile of genes expressed in human brain (Adams et al. 1991, 1992, 1993a, 1993b; Khan et al. 1992; Auffray et al. 1995), HepG2 human liver cells (Okubo et al. 1992), *Caenorhabditis elegans* (McCombie et al. 1992; Waterston et al. 1992), mouse testis (Kerr et al. 1994), and human skeletal muscle (Auffray et al. 1995).

Furthermore, this approach has been successfully applied to the identification of genes involved in genetic diseases such as the *hMLH1* gene, shown to be responsible for hereditary nonpolyposis colorectal cancer (Papadopoulos et al. 1994). Finally, the provision of coding loci would extend the database for evolutionary analysis of genome organization by permitting comparison of linkage relationships of homologous genes (O'Brien et al. 1993) and precise determination of fusion points in interchromosomal rearrangements between ancestral chromosome blocks revealed by heterologous painting between human and cattle (Hayes 1995; Solinas-Toldo et al. 1995).

To progress towards a better understanding of the mammary function, we aim to find new candidate genes that play crucial roles in the functioning and secretory activity of the mammary gland. These expressed sequences will also be used to develop the transcribed genome maps of ruminants. To this purpose, we have constructed a lactating goat mammary gland cDNA library from which nearly a thousand cDNA clones were isolated and 435 partially sequenced. The goat was chosen essentially because (i) it is

Correspondence to: F. Le Provost

The nucleotide sequence data reported in this paper have been submitted to EMBL and have been assigned the accession numbers X73542–X73548, X73704–X73804, X76318, and 271789-271924.

a representative model for ruminants and (ii) many mammary physiological studies have been carried out in this species. Here we report the identification of 77 new genes that have no similarity with sequence entries in current databases.

Materials and methods

RNA isolation. Total RNA was extracted from the mammary tissue of a freshly slaughtered lactating goat by the guanidinium thiocyanate method (Cathala et al. 1983). Poly(A)$^+$ RNA was purified from total RNA by two successive chromatographic runs on oligo(dT)-cellulose according to Aviv and Leder (1972).

cDNA library construction and screening. cDNA synthesis was performed from 2.5 μg poly(A)$^+$ RNA with the cDNA Synthesis System Plus kit (Amersham). First strand cDNA was primed with oligo(dT) and synthesized with avian myeloblastocis virus (AMV) reverse transcriptase. Second-strand synthesis was performed with *Escherichia coli* DNA polymerase I, after treatment of the mRNA/cDNA hybrid with *E. coli* ribonuclease H. Finally, the double-stranded cDNA was end-filled with T4 DNA polymerase and cloned into *Sma*I-digested and dephosphorylated pUC18. *E. coli* DH5α competent cells (BRL) were transformed to ampicillin resistance with the ligation reaction mix, according to the manufacturer's protocol. The complete lactating mammary gland cDNA library was transferred onto positively charged nylon membranes (Amersham) and hybridized with cDNA probes for the six major milk proteins (αs1-, αs2-, β-, κ-caseins, α-lactalbumin, and β-lactoglobulin), cDNA probes were radiolabeled by random priming [α-^{32}P]-dCTP (Feinberg and Vogelstein 1984). The membranes were then treated according to the recommendations of the manufacturer. Positive clones were discarded.

Northern and dot blot analysis. For Northern analysis, poly(A)$^+$ RNA (20 μg) was treated with glyoxal as described (McMaster and Carmichael 1977), separated by electrophoresis in 1% agarose gels, and transferred onto nylon membranes (Amersham). Dot blot analyses were performed as described with 20 μg of total RNA denatured in formaldehyde (6.5%). The membranes were probed as described above.

Reverse transcription. Reactions were carried out using 1–2 μg of total RNA in the presence of 50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 2 mM DTT, 2 mM of each dNTP and 40 units RNasin. The reactions were primed with 200 pM oligo(dT) by incubation for 1 h at 37°C in the presence of 200 units of Moloney murine leukemia virus (MMLV) reverse transcriptase (GIBCO/BRL-Life Technologies).

Polymerase chain reaction (PCR). In vitro DNA amplification was performed with the thermostable DNA polymerase of *Thermus aquaticus* (*Taq* DNA polymerase) in a thermal cycler (Perkin-Elmer Cetus), essentially as previously described by Saiki and associates (1988). The reaction mix (100 μl) consisted of 10 μl of 10× PCR buffer (500 mM KCl, 100 mM Tris-HCl, 15 mM $MgCl_2$, 1% Triton X-100, pH 9), 5 μl of 5 mM dNTPs mix, 50 pmol of each primer, 5–10 μl of DNA template, and 2.5 units of *Taq* polymerase (Promega). After an initial denaturing step (94°C for 10 min), the reaction mix was subjected to the following three-step cycle, which was repeated 30 times: denaturation for 1 min at 94°C, annealing for 2 min at 60°C, and extension for 1 min at 72°C. Five to 10 μl of each reaction mix were fractionated in an agarose TBE slab gel in the presence of ethidium bromide (0.5 μl/ml).

Sequencing. Double-stranded plasmid DNA templates were prepared according to the alkaline lysis protocol. DNA sequencing reactions were performed by the dideoxy nucleotide chain termination procedure (Sanger et al. 1977), with a Catalyst 800 Labstation (Applied Biosystems, ABI) using the Prism ready reaction cycle sequencing kits (ABI). Sequencing products were analyzed on polyacrylamide gels using an ABI 373A automated DNA sequencer and the accompanying software.

Sequence analyses. All computer analyses of nucleotide sequences were performed with the software package GCG (Devereux et al. 1984). Sequence data were edited manually to remove vector sequences and tailing sequences of low reliability. The databases used for the homology searches were the GenBank and EMBL (nucleic acid databases), PIR and SWISS-PROT (protein databases), and Prosite (protein motif database). Nucleic acid database searches were performed with the FASTA program (Pearson and Lipman 1988).

During the course of searches for similarity with amino acid sequences, ESTs were translated in all six reading frames, and each translation was compared with the protein sequence databases by use of the BLASTX program.

Assignment of ESTs loci. Mapping of bovine genes was performed by PCR, with the previously characterized bovine–chinese hamster somatic cell hybrid panel (Guérin et al. 1994). Primer pairs used for synteny group assignments were chosen and designed starting from the goat EST sequence and structural data of genes available in other species, including human and mouse. Presence or absence of bovine-specific PCR amplification products was tested on DNA extracted from the 36 hybrids. Assignment to a synteny group or a chromosome was assumed when the correlation coefficient was superior or equal to 0.70, with the statistical correlation test developed by Chevalet and Corpet (1986).

Results

Analysis of cDNA library. In total, 966 randomly selected clones from a goat mammary gland non-normalized and non-directionally cloned cDNA library were isolated and studied. The average size of inserts ranged between 1 and 1.2 kb. Screening with probes for the six major milk proteins (αs1-, αs2-, β-, κ-caseins, α-lactalbumin, and β-lactoglobulin) identified 531 cDNA clones, which account for ca. 60% of the mRNAs occurring in the lactating tissue. The remaining 435 clones were sequenced from one end, generating ESTs, which were compared with sequences present in nucleotide or protein databases. cDNA clones corresponding to ESTs that did not match with any sequence present in the databases were sequenced from the other end. Thus, 154 cDNA inserts were sequenced at both extremities. Therefore, 589 ESTs corresponding to 435 cloning entities were obtained.

Comparison of DNA sequences to databases. The results of sequence alignments between the 589 ESTs and sequences in databases are summarized in Fig. 1. Of the 435 cDNA clones sequenced, 118 were not studied further as they contained either polyA tail sequences only (15), were considered too short (6), were unsequenceable (6), or contained sequences that matched with the six major milk protein genes (91). Owing to local background in some instances, and in order to avoid elimination of false positives, only clones giving a strong hybridization signal were discarded in our substractive screening. This is very likely the reason why our hybridization screen failed to eliminate 91 clones corresponding to milk proteins encoding cDNAs. On the basis of database alignments with known genes from mammalian or evolutionarily distant organisms such as *E. coli, Drosophila melanogaster, Saccharomyces cerevisiae,* and *Neurospora crassa,* 234 cDNA clones (out of 317) were putatively identified. 213 ESTs matched to sequences present in nucleotide databases (EMBL or GenBank) corresponding to 119 distinct genes (Tables 1 and 2). 21 ESTs were found to encode proteins matching to sequences present in protein databases (Swiss-Prot or PIR) that correspond to 21 distinct genes (Table 3). Sequences of the 83 remaining clones do not match significantly with any current database entries and represent 77 previously uncharacterized transcripts. The total fraction of ESTs matching some sequences in databases was 90% considering all the randomly selected clones and was 74% after discarding the clones that contain cDNA encoding the six major milk proteins.

Owing to the random selection, several clones harboring the

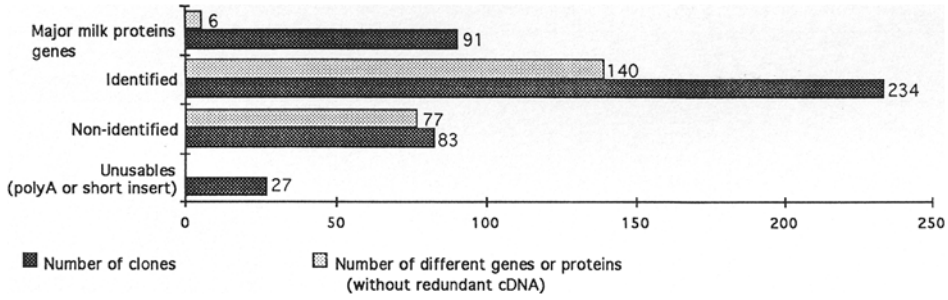

Fig. 1. Summary of results from FASTA (nucleic acid) and BLASTX (amino acid) searches for EST data which were used to assign cDNAs. Scoring criteria are described in the legends of Tables 1 and 3.

same insert were found. The rate of redundancy that corresponds to the percentage of clones containing the same cDNA, at least twice, could be calculated. We have chosen to consider two redundancy rates: whether or not cDNA clones encoding the six major milk proteins are taken into account. Considering all the clones studied, the rate of redundancy reached 82%. Within the redundant cDNA group members (767 cDNA), 622 clones consisted of only six species (major milk proteins). When these clones are substracted, the rate of redundancy falls to 44%.

Genes corresponding to the 140 distinct sequences (234 clones) identified after database searches are listed in Tables 1–3. This provides a first, although preliminary, transcriptional profile of the lactating mammary gland comprising "housekeeping" genes as well as genes specifically expressed in this tissue. Many cDNAs relevant to genes involved in protein synthesis have been found, including ribosomal proteins (L12, L27, S11, S17 . . .), acidic ribosomal phosphoprotein P1, elongation factor-1-alpha. In addition, genes known to be involved in the mammary function and expected to be expressed during lactation were identified. For example, cDNAs encoding antibacterial proteins such as lactoperoxidase (which catalyzes the peroxidation of thiocyanate) and lactoferrin (the major milk iron-binding protein) were found at one copy, whereas cDNAs encoding the glycoproteins butyrophilin and lactophorin of milk fat globule membrane (MFGM) were found in three and four copies, respectively. Regarding clones encoding lactophorin, the relevant gene has been recently characterized by Groenen and colleagues (1995) as a *GlyCAM1*-like gene specifically expressed in the lactating bovine mammary gland.

Conversely, the presence of the prokaryotic tetracycline-resistance transposon *Tn*10 right IS is rather unexpected. Indeed, 14 out of the 966 cDNA clones studied matched IS10-right with a high percentage of similarity (between 95% and 99.6%).

The presence of a Prion protein (PrP) transcript that encodes the precursor protein of scrapie (spongiform encephalography; Prusiner 1991) is also rather unexpected, even though the *PrP* gene has been shown to display a nearly ubiquitous tissue distribution (Bendheim et al. 1992). However, its expression had not been reported, up to now, in the mammary gland.

Several genes known to be expressed in mammary epithelial cells, such as acetyl coA carboxylase, lipoprotein lipase, prolactin receptor, or parathyroid hormone-like, were not identified in the mammary cDNA library. However, the presence of their transcripts has in fact been demonstrated, by RT-PCR experiments, starting from mammary tissue total RNA.

Finally, of 19 ESTs showing similarities ranging between 94% and 61%, with ESTs already deposited in databases (Table 2), two were sequences from identified genes (ribosomal proteins) and 17 were "new genes" present in human brain (13), liver (3), or fibroblast cells (2).

Comparison of expression profiles: identification of tissue-specific genes among the unidentified cDNAs. Tissue specificity and relative abundance of 18 out of the 77 unidentified clones were examined by dot blot analysis in mammary gland, liver, and brain tissues (Fig. 2). The 18 clones were selected for their typical mRNA features (poly A sequence, polyadenylation signal, open reading frame) and because the sequences of each end had been overlapped. cDNAs analyzed have been classified into five groups on the basis of their pattern of hybridization: one group comprises cDNAs from transcripts expressed at the same level in all three tissues analyzed. These genes might be involved in basic cellular functions. cDNA clones corresponding to genes predominantly expressed in the mammary gland or in the brain are gathered in two groups. Another group comprises cDNAs from transcripts predominantly found both in the mammary gland and brain, and the fifth group contains cDNAs for which no hybridization signal was observed, possibly representing very rare transcripts. Northern blot analysis of cDNAs corresponding to genes predominantly expressed in the mammary gland (Fig. 3) was used for sizing the relevant transcripts, which were found to range between 1.5 kb (clone 4-11) and 3.5 kb (clone 16-51). Two transcripts (1.3 and 1.9 kb) were identified with cDNA 18-04 as probe.

The tissue specificity of clones 4-11 and 18-04 was examined with a larger panel of tissues, by use of the RT-PCR technique. Owing to its short insert size (134 bp) and to its nucleotide sequence (many adenine or thymine repetitions), which precludes the design of oligonucleotides, clone 16-51 was not analyzed in this way. For clone 4-11, an amplified DNA fragment of the expected size (227 bp) was obtained with total RNA from the ten tissues analyzed (Fig. 4a). An amplification product was observed from liver and brain samples, where this transcript had not previously been detected by dot blot analysis. The transcript revealed by clone 18-04 was detected in all tissues except spleen and liver. However, amplification yields were considerably lower for lung, ovary, kidney, and heart than for muscle, colon, or brain (Fig. 4b). Therefore, this transcript could be considered as nearly ubiquitous, but with a large range in expression levels. However, quantitative interpretation of these data has to be considered cautiously.

Mapping cDNAs with somatic cell hybrids. Owing to the high degree of similarity observed between goat and cattle genomes (Pépin et al. 1994), goat sequences from mammary ESTs were used to design oligonucleotide primers for gene assignments in cattle, by PCR. The bovine–hamster somatic cell hybrids panel used for physical mapping of the relevant genes has been previously characterized (Guérin et al. 1994; Vaiman et al. 1994; Ma et al. 1995). The bovine chromosome assignments of syntenic groups have been determined, except for Chromosome (Chr) X, which is present in all somatic cell hybrids. Six cDNAs (lactoferrin, glucocorticoid receptor α, thioredoxin, thymosin β4, cyclophilin T, lactoperoxidase) were chosen primarily because their localization was unknown in the bovine genome. In addition, the structural organization of the relevant gene had been established in rodents or human, thus permitting positioning of primers within a single large exon or preferably within two contiguous exons separated by an intron of appropriate size for PCR. Moreover, two of these cDNAs (thioredoxin and glucocorticoid receptor α) belong to the 321 reference anchor loci, defined by O'Brien and coworkers (1993) as

Table 1. Results of FASTA searches in the GenBank and EMBL databases with EST from goat mammary gland.

Name	Species[a]	Accession number	Identity (%)	Length (bp)	Number of clones
18S ribosomal RNA	m	X00686	99.6	236	1
18S, 5,8S and 28S ribosomal RNAs	r	V01270	94.6	146	1
90-kDa heat-shock protein	h	M16660	86	386	3
acidic ribosomal phosphoprotein P1	h	M17886	91.1	394	2
actin, gamma-non muscle	o	X60733	82.6	258	1
amyloid A protein (serum)	d	M59171	78.8	321	1
amyloid-beta protein (APP)	h	M34875	65.9	214	1
antigen CD36	h	M98399	77	466	1
argininosuccinate synthetase (Ass)	m	M31700	84.8	257	1
brain-specific identifier sequence (ID) clone p1B337	r	K01677	91.3	104	1
butyrophilin	b	M35551	92.1	547	3
c-yes-1	h	M15990	71	462	1
C10 protein	b	M64923	92.1	417	1
c6.1A	h	X64643	65	206	1
caldesmon	h	M64110	82.1	464	1
calpain II regulatory subunit	b	J05065	96.9	448	2
calregulin (ERp60)	m	M92988	89.1	110	1
complement component C9	h	X02176	68.3	303	1
complete genome	v	M35027	54.8	210	1
cyclophilin, T-cell	h	V00052	94.4	408	2
cytochome P450-ssc	b	M25920	78.3	186	2
cytoskeletal gamma-actin/amy1 encoding salivary amylase	h	X04098/M18671	79.4	374	1
elongation factor 1 alpha*	o	X62245	90.1	426	2
elongation factor-1-beta	h	X60489	85.7	566	2
endothelial ligand for L-selectin (GLYCAM1)	m	M93428	54.7	601	3
fatty acid synthase	r	X13415	68.1	226	2
ferritin H chain	h	M11146	81.5	157	1
FKBP-12 protein (FKBP-12)	h	M80204	84.8	335	1
G052 protein	h	M69199	69	329	1
galactokinase (GK2)	h	M84443	84	219	1
globin gene and globin (PSI-2) pseudogene, gamma	b	M63452	83.6	535	6
globin gene and globin (PSI-3) pseudogene, beta	b	M63453	76.3	177	1
glucocorticoid receptor alpha	h	M10901	90	311	1
heterogenous nuclear RNA W16W	h	X17272	73.4	305	1
hnRNP A2 protein	h	M29065	100	190	1
Ig germline gamma-1-chain	b	X16701	90	258	1
IgJ chain	h	M12759	74	277	1
IG processed lambda chain pseudogene	b	M21660	88.3	394	2
immunoglobulin lambda light chain, germline	h	X57808	72.9	436	1
inositol triphosphate receptor (ITPR2), type 2	r	X61677	85.9	461	2
L10c5 repeat	a	X15965	70.1	562	1
lactophorin	b	D26176	93.4	624	4
lactoperoxidase (LPO)	b	M58150	83.3	210	1
lactotransferrin/lactaferrin	b	X57084	92.1	530	1
lamin A	h	M13452	85.6	118	1
mammary derived growth inhibitor	b	X51933	93.8	97	1
microglobuline, beta-2-	h	M17987	61.8	309	1
mitochondrial ATP synthase gamma subunit	b	M22463	94.8	211	2
mitochondial DNA containing repeated sequence	a	X54172	85.6	118	1
mitochondrion cytb gene for cytochrome b	g	X56289	91.6	418	1
mitochondrion, complete genome	b	J01394	84	519	21
neutrophil protein	p	M55701	96.7	123	1
ornithine decarboxylase antizyme (pseudogene)	r	D11373	75.6	397	1
phe-, val-, and leu- (5′ end) tRNA, and 12S et 16S ribosomal RNA genes	g	M55541	98.9	377	16
phospholamban	d	V00399	64.5	358	1
photolyase	n	X58713	64	236	3
polyA binding protein	h	Y00345	91	432	3
prion protein (PrP)	s	M31313	74	373	1
proliferating cell nucleolar protein P120	h	M33132	76.3	152	1
protein phosphatase 2A alpha subunit	h	M64929	90.7	463	1
prothymosin-alpha pseudogene	h	J04799	89.9	397	1
pyrroline-5-carboxylate reductase	h	M77836	86	357	1
ras-related GTP-binding protein involved in membrane traffic (rab1)	d	X56384	94.5	493	1
repetitive element					4
replacement variant histone H3,3	m	X13605	84.7	439	1
ribophorin II	h	V00282	78.3	386	1
ribosomal protein homologue L17A	h	X55954	88.2	474	1
ribosomal protein L12	r	X53504	88.2	467	1
ribosomal protein L21	r	M27905	84.6	423	1
ribosomal protein L23 (putative)	h	X53777	87.9	464	1
ribosomal protein L23a	r	X65228	87.7	398	2
ribosomal protein L27	r	X07424	87.5	392	1
ribosomal protein L27a	r	X52733	86.2	419	2
ribosomal protein L31	h	X15941	89.4	360	1

Table 1. *Continued.*

Name	Species[a]	Accession number	Identity (%)	Length (bp)	Number of clones
ribosomal protein L37	r	X66369	81.6	152	1
ribosomal protein L37a	r	X14069	86	343	1
ribosomal protein L41	h	Z12962	77.9	403	1
ribosomal protein L46	y	X01963	66	191	1
ribosomal protein L7	h	X52967	89.8	452	1
ribosomal protein S11*	h	X06617	91	277	2
ribosomal protein S17*	h	M13932	89.8	450	2
ribosomal protein S24*	h	M31520	92.2	321	3
ribosomal protein S25	h	M64716	90.1	464	4
ribosomal protein S28	r	X59277	79.1	234	1
ribosomal protein S5	r	X58465	83.6	440	1
ribosomal protein S6	h	J03537	89.2	490	1
ribosomal protein S8	r	X06423	87.2	180	1
salivary alpha-amylase	h	M18715	76.9	412	1
signal transducer CD24	h	M58664	63.7	292	1
stearoyl-CoA desaturase (SCD2)	m	M26270	70.6	235	1
succinate deshydrogenase flavoprotein subunit (SDHFP1)	b	M60879	79.7	118	1
SV40 infected cell line 14B, segment 2	r	M28630	81	526	1
thioredoxin protein	h	J04026	85.1	475	1
thymosin beta-4	h	M17733	83.6	548	1
translocation T(4:11) of ALL-1 gene to chromosome 4	h	L04731	87.1	319	1
transposable P vector conferring G418 resistance	j	X01803	94.8	252	1
transposon Tn10 DNA, 3'end		J01829	99.6	458	14
tubulin, c(beta)7	c	X07011	79.4	287	1
ubiquitin	h	M10939	89.5	440	2
ubiquitin (Uba80)	h	X63237	89.9	348	1
UDP-galactose: N-acetylglucosamine galactosyl transferase	m	J03880	75	168	1
v-fos transformation effector proteine (Fte-1)/ ribosomal protein S3a	h	M84711/M77234	90.2	417	2
valosin-containing protein (VCP)	p	M30143	86.6	367	1
ESTs	h	Table 2			

Quality of match is given as percentage identity and length in base pairs. The general criteria used for scoring a cDNA as having a significant match with a nucleotide database entry were >150 bp in length and >60% identity. In a few instances these criteria were relaxed when a significant degree of relatedness was evident. When multiple significant similarities were found for a single cDNA, only the highest scoring hit is included.

[a] Species abbreviations used are as follows: a, rabbit, b, bovine; c, chicken; d, dog; e, *Escherischia coli;* g, goat; h, human; j, *Drosophila melanogaster;* k, *Ascaris suum;* m, mouse; n, *Neurospora crassa;* o, rabbit; p, pig, r, rat; s, sheep; v, virus; y, yeast. *: matches with the same cDNA from several organisms.

Table 2. Matches with the EST data set.

Clone number (size)	Matching EST (size)	Tissue	Accession number	Identity (%)	Length (bp)	Number of clones
1.09 (530)	HUM000AA21 (452)	liver	D11489	73.5	151	1
2.09 (259)	HUM0S12E01 (140)	liver	D11977	80.7	140	1
2.51 (243)	T03559 (447)	brain	T03559	71.8	255	2
3.35 (351)	HUMXT02070 (319)	brain	M85554	83.8	167	1
4.09 (396)	MMTSG153X (263)	brain	X61808	86.2	130	1
4.31 (287)	HUMXT01030 (285)	brain	M78882	80.3	290	1
6.10 (546)	HSAAADLOR (190)	brain	Z21073	80.2	162	2
7.19 (176)	HUMXT01459 (387) (ribosomal protein YL10)	brain	M77875	83.7	178	1
11.31 (359)	HUMXT02002 (381)	brain	M85486	60.9	289	1
11.42 (410)	HUMXT00691 (366)	brain	M78543	71.2	212	2
14.52 (356)	HUMHM01E12 (111)	liver	D11779	88.1	109	1
15.24 (327)	HSAFIF031 (301)	fibroblast	Z21890	94.1	303	2
15.50 (436)	HUMXT00065 (183)	brain	M62009	78.3	157	1
16.03 (424)	HUMXT01627 (308) (ribosomal protein L1a)	brain	M78040	71.4	199	1
19.08 (504)	HUMXT00123 (368)	brain	M62066	76.2	341	1

By subjecting the 435 sequences reported in this paper to a similarity search with ESTs present in the database, overlaps were found. Clone names with sizes (in base pair) of the ESTs are listed in the first column. In the second column is given the locus name of the matching EST as well as its size. Columns five and six show the identity (%) and the length of the overlap. Clones 7-19 and 16-03 match with ESTs which are similar to ribosomal protein YL10 and ribosomal protein L1a, respectively.

suitable markers for comparative gene mapping in mammals. Results of these mapping experiments are given in Table 4. It is worth noting that correlation coefficients are between 0.83 and 1.00, and that the six loci are assigned to five different synteny groups (U12, U22, U18, U20, U21) identifying five different chromosomes (7, 8, 19, 22, 23) and to Chr X. Mapping of lactoferrin and lactoperoxidase by fluorescent in situ hybridization (Schwerin et al. 1994; Hayes et al. 1993) confirmed our assignments of Chr 19 and 22 to synteny groups U12 and U21, respectively.

Discussion

To our knowledge, this study represents the first systematic sequence analysis of an mRNA population occurring in the lactating

Table 3. Results of ESTs similarities in the PIR and Swiss-PROT databases.

Putative identification	Species	Accession number	Length (AA)	Identity (%)	Probability
40S ribosomal protein S4, X isoform	h	RS4X_HUMAN	14	100	$7.4\ 10^{-3}$
60S ribosomal protein L30A	y	RL3A_YEAST	108	50	$2.4\ 10^{-35}$
60S ribosomal protein L39	r	RL39_RAT	50	98	$3.4\ 10^{-38}$
ATP-dependent CLP protease proteolytic subunit	e	CLPP_ECOLI	46	45	$1.8\ 10^{-8}$
highly repeated long interspersed DNA family	r	B25556	36	44	$3.2\ 10^{-10}$
DNA-binding protein TAXREB 67 (transcription factor ATF-4)	h	ATF4_HUMAN	59	96	$8.7\ 10^{-43}$
G2/mitotic-specific cyclin 2	y	CG22_YEAST	56	44	$3.7\ 10^{-5}$
galactokinase	y	GAL1_KLULA	42	42	$1.4\ 10^{-10}$
hepatic lectin homolog	v	LEC2_FOWPM	58	34	$2.4\ 10^{-11}$
hypothetical 45.0 kDa protein in MBP1-HMR intergenic region	y	YCY4_YEAST	18	72	$3.8\ 10^{-10}$
hypothetical protein in ubig	e	YFAC_ECOLI	40	67	$6.2\ 10^{-14}$
L1 elements	h	B34087	36	41	$2.8\ 10^{-4}$
line-1 reverse transcriptase factor	h	LIN1_HUMAN	23	69	$3.8\ 10^{-7}$
low affinity immunoglobulin epsilon FC receptor	h	FCE2_HUMAN	40	35	$4.6\ 10^{-7}$
lysosomal acid phosphatase precursor	r	PPAL_RAT	39	46	$7.4\ 10^{-6}$
NADH dehydrogenase	j	S01186	25	52	$9.2\ 10^{-3}$
penicillin-binding protein 1A	e	PBPA_ECOLI	83	49	$5.7\ 10^{-20}$
procollagen alpha 2(IV) chain precursor	k	CA24_ASCSU	30	40	$4.4\ 10^{-5}$
QM protein	h	QM_HUMAN	100	95	$1.2\ 10^{-73}$
ras-related protein rab-2	h	RAB2_HUMAN	34	47	$1.5\ 10^{-4}$
tum-transplantation antigene P198	m	JL0149	131	95	$3.8\ 10^{-98}$

Protein database searches were conducted with a program for peptide comparisons (BLASTX). Sequence similarities identified were considered statistically significant with a Poisson p-value < 0.01. The Poisson p-value is the probability of a high score occurring by chance, taking into account the number of residues in the query sequence and the database. The general criteria used for scoring a cDNA as having a significant match with a nucleotide database entry were: 30 amino acid residues in length and >40% identity. In a few instances these criteria were relaxed when a significant degree of relatedness was evident. When multiple significant similarities were found for a single cDNA, only the highest scoring hit is included. Species abbreviations used are as for Table 1.

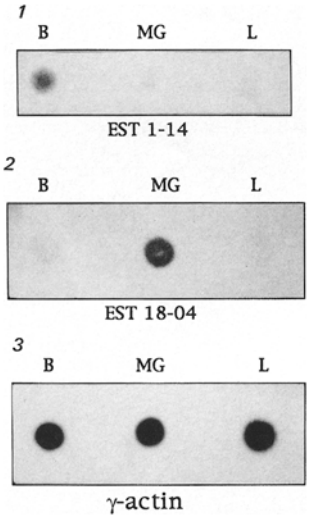

Fig. 2. Analysis of the expression of unknown cDNAs in goat mammary gland (MG), brain (B), and liver (L) by dot analysis. Total RNA (20 μg) was isolated, immobilized on nylon membranes, and hybridized with labeled cDNA inserts. The cDNA inserts could be classified into five groups based on their hybridization pattern, and representative clones from cDNAs corresponding to genes predominantly expressed in the brain (EST1-14) or in the mammary gland (EST18-04) were chosen to illustrate the patterns of expression observed. Panel **1:** 1-14; panel **2:** 18-04; panel **3:** γ-actin: hybridization control.

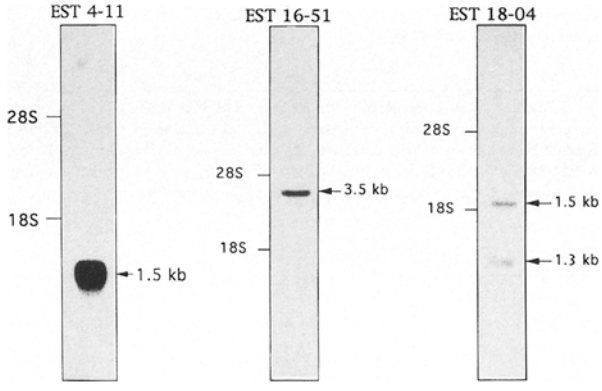

Fig. 3. Northern blot analysis of three cDNAs that are expressed predominantly in the mammary gland. Poly A^+ RNAs (20 μg) from goat mammary gland were separated on 1% agarose gels. After transfer, each blot was probed with cDNA 4-11, 16-51, 18-04; the sizes of the mRNAs detected by these clones are approximately 1.5, 3.5, and 1.3 and 1.9 kb, respectively. Positions of 28S and 18S ribosomal RNAs are indicated on the left.

mammary gland of any species. It provides a survey of the genes expressed in this tissue and gives an insight into the transcriptional activity of the goat mammary gland during lactation.

In total, 589 ESTs were generated from 435 selected cDNA clones. After discarding the cDNAs encoding the six major milk proteins, examination of these ESTs for similarities with sequence databases identified 234 cDNAs corresponding to 140 unique genes or proteins. The 83 remaining clones, which are not significant matches to any current database entries, represent 77 sequences unrelated to previously described genes. Since "housekeeping" genes have been widely studied and characterized in human, the new cDNAs reported here possibly encode genes involved in functions specific for the mammary tissue. However, dot blot as well as RT-PCR experiments have shown that these transcripts are found not only in the mammary gland; thus, the relevant genes are probably expressed preferentially, but not exclusively, in the mammary tissue. Nevertheless, liver and brain functions are different from those of the mammary gland, which is a secretory tissue. Consequently, cDNAs expressed predominantly in the mammary gland (clones 4-11 and 18-04) remain good candidates in our search for genes fulfilling a basic role in the mammary function.

Advantages and drawbacks of methods used to construct and characterize the cDNA library. For this study, a non-normalized cDNA library had been constructed. In some cases, various forms of prescreening or normalization have been proposed to construct

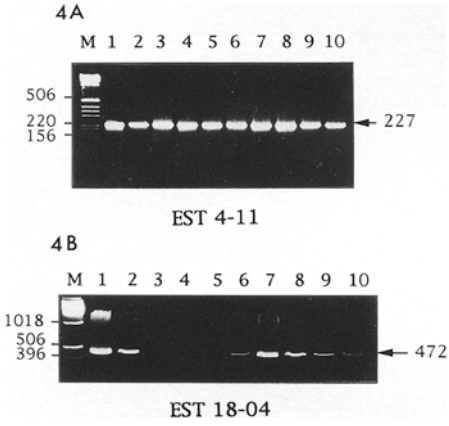

Fig. 4. Detection of 4-11 (4A) and 18-04 (4B) transcripts from different goat tissues, after agarose gel electrophoresis of products amplified by RT-PCR. Total RNA (5 μg) was extracted from mammary gland (1), brain (2), liver (3), spleen (4), lung (5), ovary (6), muscle (7), colon (8), kidney (9), heart (10). Sizes (in bp) of amplified DNA fragments are given on the right, and those of molecular size markers (lane M) on the left.

cDNA libraries containing equal amounts of cDNAs for genes expressed in a given cell, tissue, or organ (Ko 1990; Patanjali et al. 1991). Indeed, in mammalian cells, given the expression level of the relevant genes, mRNAs have been classified in high, middle, and low abundance classes (Hastie and Bishop 1976). The non-normalized library construction had been chosen, first because experiments carried out with a human brain cDNA library (Adams et al. 1991, 1992) showed that the majority of the ESTs generated without normalization were represented only once. Second, probes required for the elimination of cDNAs corresponding to highly abundant mRNAs were available in the laboratory. Whereas this mRNA class does not usually exceed 20% of the total messengers, in the mammary gland it accounts for at least 60%, which only encode six different proteins (αs1-, αs2-, βs, κ-caseins, α-lactalbumin, and β-lactoglobulin). In this study, the cDNAs encoding these six main milk proteins comprised 66% of the clones analyzed. This percentage being in agreement with the expected value, one can assume that representation of the mammary mRNAs population, in our library, is not biased.

The rate of redundancy, calculated for our goat mammary cDNAs library, is significantly higher (44%) than that reported by McCombie and associates (1992) for a *C. elegans* cDNA library (24%) and considerably higher than that (less than 5%) obtained for the human brain cDNA libraries (Adams et al. 1991, 1992). The high rate of redundancy observed here suggests that the middle abundant mRNA class constitutes a significantly larger fraction of total mRNAs than in the human brain. Therefore, we can expect that only a few genes of the low abundant class are present in the library that we have constructed.

However, the redundancy of uncharacterized transcripts is not easy to estimate, since, owing to the way the sequencing was performed, two ESTs differing in their 5′ end can be derived from non-overlapping sequences of the same gene. However, 3′ sequencing provides a more constant reference point for clone-to-clone comparisons, and some structural features, such as a polyadenylation signal (AATAAA) upstream from a polyA tail, indicate an actual mRNA origin for the cloned cDNA. This feature has permitted the detection of six clones found twice among the 83 unidentified clones.

Identification of sequenced clones. Single-run DNA sequencing has proven to be an efficient method to obtain preliminary data on cDNA inserts. Our results confirm that sufficient information is contained in a 150–400 nucleotide sequence from one sequencing run for preliminary identification of the cDNA. However, as each cDNA was sequenced only once, some errors (deletion, insertion, base miscalling) may have occurred. This kind of error does not really affect the percentage of homology between two nucleotide sequences, even though it can shift the open reading frames and change the deduced amino acid sequences. However, this may have awkward consequences, especially when amino acid sequence comparison as well as subsequent search for protein domains are addressed.

After substraction of the major milk protein mRNAs, the number of ESTs displaying significant similarities with sequences present in databases (74%) is higher than that obtained from human brain (17%; Adams et al. 1992) and, to a less extent, from *C. elegans* (42%; McCombie et al. 1992) cDNA libraries. Even though the three libraries were non-normalized, owing to the nature of the starting biological material, one could expect such differences. Indeed, brain is considered as the tissue in which the number of expressed genes to be discovered is highest. The cDNA library from *C. elegans* was constructed with mRNAs extracted from the whole organism and should, therefore, reflect a large diversity of transcripts. Conversely, the mammary gland is a very specialized organ involving fewer genes. Moreover, since 1992, databases have been largely supplied with new sequences of human and rodent origins. However, the percentage of sequence identities between goat and these species is generally high enough to allow identification of cDNAs, despite the fact that a large number of sequences span 3′UTR, which are usually less well conserved than coding regions.

cDNAs identified from alignment analysis with sequences in the databases could be divided into functional categories covering the biological spectrum; thus, a phenotypic characterization of the tissue begins to emerge. Several categories (cytoskeleton protein, nuclear protein, membrane protein, secretory protein) are found in mammary epithelial cells and in all cell types. Nevertheless, given the function of the mammary gland, which synthesizes and secretes large amounts of protein during lactation, it is reassuring to find a large number (32) of transcripts from genes encoding proteins and enzymes involved in milk protein synthesis. The analysis of the cDNA sequences generated consistently identified transcripts from genes encoding minor milk proteins such as lactoperoxidase, lactoferrin, or proteins from the fat globule membranes (lactophorin and butyrophilin).

With large-scale cDNA-based approaches, DNA contaminants may be inadvertently cloned during the construction of the library. Up to now, this is a hypothesis concerning the unexpected occurrence of prokaryotic IS10-right clones in our goat mammary gland cDNA library. Transposon *Tn*10, which is widespread among Gram-negative bacteria, is a compound transposon composed of two copies of an IS element flanking a drug resistance gene encoding a tetracycline-inducible, membrane-located peptide that confers resistance to the antibiotic by promoting its efflux from the bacterial cell. The presence of transcripts encoding IS10-right in mammary tissue of goat could reflect an infective status of the udder, owing to mastitis-related Gram-negative pathogens, or a post-RNA extraction bacterial contamination or the occurrence of IS10-right sequences in the goat genome transcribed in the mammary tissue. Two experimental arguments tend to refute the bacterial origin. First, Northern blot analyses (data not shown) revealed IS10-right transcripts among RNA extracted from the mammary gland of seven lactating goats not displaying any clinical mastitis. Second, if the contamination hypothesis had to retained, one could have expected that other prokaryotic transcripts might also have been detected among clones sequenced. Likewise, the mammalian origin of this transcript remains highly questionable, since Southern blot analyses did not succeed in detecting any trace of IS10-right sequence at the goat genomic DNA level. However, direct DNA transfer from intracellular pathogenic bacteria to mammalian cells has been recently demonstrated (Courvalin et al.

Table 4. Localization of cDNA sequences with bovine-hamster cell hybrids.

Name	Primers	Synteny group	Bovine chromosome	Correlation coefficient	Human chromosome
Lactoferrin* (LTF)	GCTTCCAGCTCTTTGGCTCCA TGAGGTTCTTCAAGGCGGTCAAG GCCATTGTCAAGAGGGAAGGA	U12	22	0.84	3
Glucocorticoid receptor alpha (GRL)	TCCAAAAATGTTTGGAAGCAATAG GCAGAGTGTGAAGTCAAATGC	U22	7	0.88	5
Thioredoxin (TXN)	CCTTATTAGCTCCAGAAAATTCAC CCGATATGGCTGAGATTGAGAAG	U18	8	1.00	9
Thymosin beta4	ATGTACAGTGCATACTGGCGGTG CCTGCTTTCACAGAATAATTCCAG		X	0.87	nl[b]
Cyclophilin T (CPHM)	CATTTGCCATGGACAAGATGCCAG ATTGTGCTAGGTAGTGAGATGCAG	U20	23	0.83	nl
Lactoperoxidase[a] (LPO)	GTCTTTGATTATCCTCCAGGTGTTG	U21	19	1.00	7

Primer sequences are listed 5′ to 3′ reading left to right. Human localization data originate from GDB (Genome DataBase).
[a] Results that are described elsewhere: Le Provost et al. (1994b) and Cals et al. (1994).
[b] nl, not localized.

1995). A somatic delivery process of bacterial integrative (?) DNA, directed toward mammary epithelial cells, stably inherited and expressed by the cell progeny, might have occurred, leading to the expression in the mammary gland of IS10-right encoding transcripts. We are currently investigating such a hypothesis.

The Prion protein (PrP) transcript, usually detected in the brain but also in a wide variety of other tissues and cell types, has been detected in the mammary tissue. PrP is considered to be the agent of scrapie, a transmissible degenerative disease of the central nervous system occurring in sheep and goat (for a review, see Prusiner 1991). Since it has been shown that allelic forms at the PrP locus might be a predisposition to spongiform encephalopathies (Laplanche et al. 1993), milk might be a transmission medium for passage of scrapie from the mother to the young, even though the infectiousness of milk is usually considered as weak (Dickinson et al. 1974).

Whereas some genes were not expected to be expressed in the mammary tissue, transcripts from others expected to be rather well represented were lacking within our cDNA library. This situation is exemplified by acetyl coA carboxylase and lipoprotein lipase, which were detected only by RT-PCR experiments (Leroux, personal communication).

Regarding the prolactin gene, which has been shown recently to be ectopically transcribed in mammary epithelial cells (Le Provost et al. 1994a), no relevant cDNA clone was found among the thousand clones examined, thus confirming its very low level of expression in this tissue as compared with that recorded in the pituitary gland.

Contribution to ruminants genome mapping effort. The EST approach used here represents a general strategy to accelerate the development of bovine genome mapping programs by allowing, with a PCR-based method, a rapid assignment of genes to specific chromosomes, as proposed by Sikela and coworkers (Wilcox et al. 1991). This "expression map" strategy is a complementary approach to genetic mapping, which is mainly based on the positioning of non-coding DNA polymorphisms, essentially multi-allelic microsatellite-type markers (Eggen and Fries 1995).

Genes encoding thymosin β4 and cyclophilin T have been assigned to bovine Chrs X and 23 (*BTAX* and *BTA23, Bos taurus*), respectively. Assignment of these genes in human is unknown. Since genes belonging to the synteny group corresponding to Chrs X and 6 in human (HSAX and HSA6, *Homo sapiens*) are found syntenic in cattle and have been mapped to *BTAX* and *BTA23* respectively, one can expect cyclophilin T and thymosin β4 to be localized on HSA6 and HSAX, respectively. Regarding the lactoferrin gene (LTF), its localization in human (Teng et al. 1987) led us to suppose a physical mapping on *BTA1,* whereas Schwerin et al. (1994) have localized it by FISH on *BTA22.* In fact, since serum transferrin (*STF*) had been previously mapped on *BTA1, LTF* and *STF* loci are not localized on the same bovine chromosome, in contrast to the situation observed in human and mice (Le Provost et al. 1994b). These results are very informative in terms of evolutionary relationships and will help to reach a better understanding of the hypothetical rearrangements having occurred in the ancestral chromosomes. Elsewhere, we confirm at the gene level the correspondences between *BTA7, 8,* and *19,* and HSA5, 9, and 17, respectively, shown by chromosome painting (Hayes 1995; Solinas-Toldo et al. 1995). This is exemplified by the lactoperoxidase gene, localized on *BTA19* entirely and only painted by HSA17, which is shown to carry the lactoperoxidase locus (Hayes, personal communication). Likewise, the glucocorticoid receptor, which was shown to belong to the synteny group U 22, is assigned to *BTA7,* which is painted by HSA19 and HSA5. Localization of human genes on HSA5 (Francke and Foellmer 1989) strongly supports such a result. Finally, as far as the gene encoding thioredoxin is concerned, its assignment to BTA8 confirms the correspondence between HSA9 and *BTA8.* Additional support for this proposal has been recently provided by Heppell-Parton and colleagues (1995), who have firmly established the chromosomal location of the human thioredoxin gene to HSA9q31.

A well-detailed gene map may provide the basis for a unified approach to comparative analysis of mammalian species genomes and help to dissect the evolution of genome organization (O'Brien et al. 1993). Moreover, it will define more precisely the distribution of evolutionarily conserved segments between bovine and human karyotypes. In this connection, we are currently carrying on chromosomal assignment, using somatic cell hybrids, as well as fluorescent in situ hybridization, of novel ESTs and ESTs corresponding to genes of unknown localization, in the bovine genome.

Regarding our search for genes playing a crucial role in the mammary function, we are characterizing further unknown genes preferentially expressed in the mammary tissue. In addition, we have started to explore the differential display approach, recently described by Liang and Pardee (1992), to identify and isolate differentially expressed genes.

Acknowledgments. We are greatly indebted to Michèle Nocart and Gérard Guérin for performing somatic cell mapping. We thank François Grosclaude for his constant interest in this work, and Charles Auffray and his colleagues for helpful discussions. We also thank Phil L'Huillier and Hélène Hayes for correcting the manuscript. This work was financially supported in part by a government grant from the Groupement de Recherches et d'Etudes sur les Génomes (GREG).

References

Adams, M.D., Kelley, J.M., Gocayne, J.D, Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olden, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R., Venter, J.C. (1991). Complementary

DNA sequencing: expressed sequence tags and human genome projet. Science 252, 1651–1656.

Adams, M., Dubnick, M., Kerlavage, A., Moreno, R., Kelley, J., Utterback, T., Nagle, J., Fields, C., Venter, J. (1992). Sequence identification of 2,375 human brain genes. Nature 355, 632–634.

Adams, M., Bento Soares, M., Kerlavage, A., Fields, C., Venter, J. (1993a). Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. Nature Genet. 4, 373–380.

Adams, M.D., Kerlavage, A.R., Fields, C., Venter, J.C. (1993b). 3,400 new expressed sequence tags identify diversity of transcripts in human brain. Nature Genet. 4, 256–267.

Auffray, C., Behar, G., Bois, F., Bouchier, C., Da Silva, C., Devignes, M.-D., Duprat, S., Houlgate, R., Jumeau, M.-N., Lamy, B., Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pietu, G., Pouliot, Y., Sebastiani-Kabatchis, C. (1995). IMAGE: intégration au niveau moléculaire de l'analyse du génome humain et de son expression. C. R. Acad. Sci. 318, 263–272.

Aviv, H., Leder, P. (1972). Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA 69, 1408–1412.

Beckmann, J., Soller, M. (1988). Detection of linkage between marker loci and loci affecting quantitative traits in crosses between segregating populations. Theor. Appl. Genet. 76, 228–236.

Bendheim, P.E., Brown, H.R., Rudelli, R.D., Scala, L.J., Goller, N.L., Wen, G.Y., Kascsak, R.J., Cashman, N.R., Bolton, D.C. (1992). Nearly ubiquitous tissue distribution of the scrapie agent precursor protein. Neurology 42, 149–156.

Cals, M.-M., Guillomot, M., Martin, P. (1994). The gene encoding lactoperoxidase is expressed in epithelial cells of the goat lactating mammary gland. Cell. Mol. Biol. 40, 1143–1150.

Cathala, G., Savouret, J.-F., Mendez, B., West, B.L., Karin, M., Martial, J.A., Baxter, J.D. (1983). A method for isolation of intact, translationally active ribonucleic acid. DNA 2, 329–335.

Chevalet, C., Corpet, F. (1986). Statistical decision rule concerning synteny or independence between markers. Cytogenet. Cell Genet. 43, 132–139.

Courvalin, P., Goussard, S., Grillot-Courvalin, C. (1995). Gene transfer from bacteria to mammalian cells. C. R. Acad. Sci. 318, 1207–1212.

Devereux, J., Haeberli, P., Smithies, O. (1984). A competitive set of sequence analysis programmes for the VAX. Nucleic Acids Res. 12, 387–385.

Dickinson, A., Stamp, J., Renwick, C. (1974). Maternal and lateral transmission of scrapie in sheep. J. Comp. Pathol. 84, 19–25.

Eggen, A., Fries, R. (1995). An integrated cytogenetic and meiotic map of the bovine genome. Anim. Genet. 26, 215–236.

Feinberg, A., Vogelstein, B. (1984). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 137, 266–267.

Francke, U., Foellmer, B. (1989). The glucocorticoid receptor gene is in 5q31-q32. Genomics 4, 610–612.

Geldermann, H. (1975). Investigations on inheritance of quantitative characters in animals by gene markers; I. Methods. Theor. Appl. Genet. 46, 319–330.

Gellin, J., Grosclaude, F. (1991). Analyse du génome des espèces d'élevage. INRA Prod. Anim. 4, 97–105.

Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A., Sargeant, L., Sorensen, A., Steele, M., Zhao, X., Womack, J., Hoeschele, I. (1995). Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 139, 907–920.

Groenen, M.A.M., Dijkhof, R.J.M., van der Poel, J.J. (1995). Characterization of a *GlyCAM1*-like gene (glycosylation-dependent cell adhesion molecule 1) which is highly and specifically expressed in the lactating bovine mammary gland. Gene 158, 189–195.

Grosclaude, F., Ricordeau, G., Martin, P., Remeuf, F., Vassal, L., Bouillon, J. (1994). Du gène au fromage: le polymorphisme de la caséine α_{s1} caprine, ses effets, son évolution. INRA Prod. Anim. 7, 3–19.

Guérin, G., Eggen, A., Vaiman, D., Nocart, M., Laurent, P., Béchet, D., Ferrara, M. (1994). Further characterisation of a somatic cell hybrid panel. Ten new assignments to the bovine genome. Anim. Genet. 25, 31–35.

Hastie, N.D., Bishop, J.O. (1976). The expression of three abundance classes of messenger RNA in mouse tissues. Cell 9, 761–774.

Hayes, H. (1995). Chromosome painting with human chromosome-specific DNA libraries reveals the extent and distribution of conserved segments in bovine chromosomes. Cytogenet. Cell Genet. 71, 168–174.

Hayes, H., Popescu, P., Dutrillaux, B. (1993). Comparative gene mapping of lactoperoxidase, retinoblastoma, and α-lactalbumin genes in cattle, sheep, and goats. Mamm. Genome 4, 593–597.

Heppell-Parton, A., Cahn, A., Bench, A., Lowe, N., Lehrach, H., Zehetner, G., Rabbitts, P. (1995). Thioredoxin, a mediator of growth inhibition, maps to 9q31. Genomics 26, 379–381.

Kerr, S.M., Vambrie, S., McKay, S.J., Cooke, H.J. (1994). Analysis of cDNA sequences from mouse testis. Mamm. Genome 5, 557–565.

Khan, A., Wilcox, A., Polymeropoulos, M., Hopkins, J., Stevens, T., Robinson, M., Orpana, A., Sikela, J. (1992). Single pass sequencing and physical and genetic mapping of human brain cDNAs. Nature Genet. 2, 180–185.

Ko, M. (1990). An 'equalized cDNA library' by the reassociation of short double-stranded cDNAs. Nucleic Acids Res. 18, 5705–5711.

Laplanche, J., Chatelain, J., Westaway, D., Thomas, S., Dussaucy, M., Brugere-Picoux, J., Launay, J. (1993). PrP polymorphisms associated with natural scrapie discovered by denaturing gradient gel electrophoresis. Genomics 15, 30–37.

Le Provost, F., Leroux, C., Martin, P., Gaye, P., Djiane, J. (1994a). Prolactin gene expression in ovine and caprine mammary gland. Neuroendocrinology 60, 305–313.

Le Provost, F., Nocart, M., Guérin, G., Martin, P. (1994b). Characterization of the goat lactoferrin cDNA: assignment of the relevant locus to bovine U12 synteny group. Biochem. Biophys. Res. Commun. 203, 1324–1332.

Liang, P., Pardee, A.B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 257, 967–971.

Ma, R., Beever, J., Da, Y., Green, C., Russ, I., Park, C., Heyen, D., Everts, R., Fisher, S., Overton, K., Teale, A., Kemp, S., Hines, H., Guérin, G., Lewin, H. (1995). A male linkage map of the cattle (*Bos taurus*) genome. J. Hered., submitted.

McCombie, W., Adams, M., Kelley, J., FitzGerald, M., Utterback, T., Khan, M., Dubnick, M., Kerlavage, A.R., Venter, J.C., Fields, C. (1992). *Caenorhabitis elegans* expressed sequence tags identify gene families and potential disease gene homologues. Nature Genet. 1, 124–131.

McMaster, G., Carmichael, G.G. (1977). Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. Proc. Natl. Acad. Sci. USA 74, 4835–4838.

O'Brien, S.J., Womack, J.E., Lyons, L.A., Moore, K.J., Jenkins, N.A., Copeland, N.G. (1993). Anchored reference loci for comparative genome mapping in mammals. Nature Genet. 3, 103–112.

Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y., Matsubara, K. (1992). Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. Nature Genet., 2, 173–179.

Papadopoulos, N., Nicolaides, N.C., Wei, Y.-F., Ruben, S.M., Carter, K.C., Rosen, C.A., Haseltine, W.A., Fleischmann, R.D., Fraser, C.M., Adams, M.D., Venter, J.C., Hamilton, S.R., Petersen, G.M., Watson, P., Lynch, H.T., Peltomäki, P., Mecklin, J.-P., de la Chapelle, A., Kinzler, K.W., Vogelstein, B. (1994). Mutation of a mutL homolog in hereditary colon cancer. Science 263, 1625–1629.

Patanjali, S.R., Parimoo, S., Weissman, S.M. (1991). Construction of a uniform-abundance (normalized) cDNA library. Proc. Natl. Acad. Sci. USA 88, 1943–1947.

Pearson, W., Lipman, D. (1988). Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA 85, 2444–2448.

Pépin, L., Amigues, Y., Lépingle, A., Berthier, J.-L., Bensaïd, A., Vaiman, D. (1994). Sequence conservation of microsatellites between *Bos taurus* (cattle), *Capra hircus* (goat) and related species. Examples of use in parentage testing and phylogeny analysis. Heredity 74, 53–61.

Prusiner, S.B. (1991). Molecular biology of prion diseases. Science 252, 1515–1522.

Saiki, R., Gelfand, D., Stoffel, S., Scharf, S., Higuchi, R., Horn, G., Mullis, K., Erlich, H. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239, 487–491.

Sanger, F., Nicklen, S., Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463–5467.

Schwerin, M., Solinas Toldo, S., Eggen, A., Brunner, R., Seyfert, H.M., Fries, R. (1994). The bovine lactoferrin gene (LTF) maps to chromosome 22 and syntenic group U12. Mamm. Genome 5, 486–489.

Solinas-Toldo, S., Lengauer, C., Fries, R. (1995). Comparative genome map of human and cattle. Genomics 27, 489–496.

Stranzinger, G. (1987). Gene mapping and linkage. Anim. Genet. 18, 111–116.

Teng, C., Pentecost, B., Marshall, A., Solomon, A., Bowman, B., Lalley, P., Naylor, S. (1987). Assignment of the lactotransferrin gene to human chromosome 3 and to mouse chromosome 9. Somatic Cell Genet. 13, 689–693.

Vaiman, D., Mercier, D., Moazami-Goudarzi, K., Eggen, A., Ciampolini, R., Lépingle, A., Velmala, R., Kaukinen, J., Varvio, S.L., Martin, P., Levéziel, H., Guérin, G. (1994). A set of 99 cattle microsatellites: characterization, synteny mapping, and polymorphism. Mamm. Genome 5, 288–297.

Waterston, R., Martin, C., Craxton, M., Huynh, C., Coulson, A., Hillier, L., Durbin, R., Green, P., Shownkeen, R., Halloran, N., Metzstein, M., Hawkins, T., Wilson, R., Berks, M., Du, Z., Thomas, K., Thierry-Mieg, J., Sulston, J. (1992). A survey of expressed genes in *Caenorhabditis elegans*. Nature Genet. 1, 114–123.

Wilcox, A., Khan, A., Hopkins, J., Sikela, J. (1991). Use of 3′ untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs: implications for an expression map of the genome. Nucleic Acids Res. 19, 1837–1843.